

CLAIMS

What is claimed is:

1 1. A method of searching images for data contained within said images, said
2 method comprising the steps of:

3 providing a plurality of training images;

4 extracting image attributes from said training images;

5 classifying training images according to said extracted attributes;

6 selecting a particular classifier for each group of training images;

7 collecting a plurality of images available from remotely connected
8 computers;

9 indexing said collected images; and

10 providing an index of said collected images for interrogation by users.

1 2. A method as in claim 1 wherein the step of extracting attributes comprises
2 identifying image features characteristic of a particular numerically generated image
3 type.

1 3. A method as in claim 2 wherein the training images include groups of one or
2 more of charts selected from a plurality of known charts consisting of column charts,
3 bar charts, line charts, pie charts, scatter charts, area charts, surface charts, and three-
4 dimensional charts.

1 4. A method as in claim 3 wherein the step of extracting attributes identifies image
2 features and assigns any identified image features as attributes, said image features
3 including horizontal lines, vertical lines, percentage of white area, circular arcs and text.

1 5. A method as in claim 1 wherein the step of classifying images comprises
2 providing image attributes for each image to a plurality of classifiers, said plurality of
3 classifiers being selected from the group consisting of classification trees, discriminant
4 functions, regression trees, support vector machines, neural nets and hidden Markov
5 models.

1 6. A method as in claim 5 wherein one of said classifiers is selected for each chart
2 type.

1 7. A method as in claim 5 wherein said collected images are indexed in the
2 indexing step by extracting attributes and providing extracted attributes to said
3 classifiers, said classifiers identifying an image type, said images being indexed
4 according to said attributes and said image type.

1 8. A method of extracting data from images located on remotely connected
2 machines, said images having been indexed according to claim 7, said method
3 comprising the steps of:
4 receiving a query from a user;
5 identifying indexed images responsive to said query; and
6 presenting identified images to said user in response to said query.

1 9. A method as in claim 8 wherein prior to presenting identified images to the user,
2 chart data is extracted from the images and the extracted chart data is presented to the
3 user, the user being allowed to select images for viewing.

1 10. A method as in claim 9 further comprising the steps of:
2 providing additional data as supplemental data to said extracted chart data;
3 and
4 generating a chart representative of supplemented data, said generated chart
5 being provided responsive to said query.

1 11. A search engine for searching images located on remotely connected machines
2 and extracting data from said images, said search engine comprising:
3 means for receiving a query from a user;
4 means for identifying images responsive to said query; and
5 means for presenting identified images to said user in response to said query.

1 12. A search engine as in claim 11 wherein the means for identifying images
2 comprises a plurality of classifiers, said plurality of classifiers being selected as
3 optimally identifying a particular image type and selected classifiers being selected from
4 the group consisting of classification trees, discriminant functions, regression trees,
5 support vector machines, neural nets and hidden Markov models.

1 13. A search engine as in claim 12 further comprising data extraction means for
2 extracting chart data from images, the extracted chart data being presented to the user
3 for selecting images for viewing.

1 14. A search engine as in claim 13 further comprising:
2 means for supplementing said extracted chart data with supplemental data;
3 and
4 chart generation means for generating a chart representative of supplemented
5 chart data, said generated chart being provided responsive to said
6 query.

1 15. A search engine as in claim 14 wherein the image types include column charts,
2 bar charts, line charts, pie charts, scatter charts, area charts, surface charts, and three-
3 dimensional charts.

1 16. A search engine as in claim 15 wherein image features including horizontal
2 lines, vertical lines, percentage of white area, circular arcs and text.

1 17. A search engine as in claim 16 further including an image identification trainer
2 comprising:

- 3 means for providing a plurality of pre-classified training images;
4 means for classifying training images according to extracted image features;
5 means for selecting a particular one of said plurality of classifiers as being an
6 optimum for classifier for a corresponding group of training images;
7 means for collecting a plurality of images available from remotely connected
8 computers;
9 means for indexing said collected images;
10 means for storing an index of indexed said collected images; and
11 means for providing said index for interrogation by users.

1 18. A computer program product fixed in a computer useable medium for searching
2 images located on remotely connected machines and extracting data from said images,
3 said computer program product comprising:

- 4 computer code means for receiving a query from a user;
5 computer code means for identifying images responsive to said query; and
6 computer code means for presenting identified images to said user in
7 response to said query.

1 19. A computer program product as in claim 18 wherein the computer code means
2 for identifying images comprises a plurality of classifiers.

1 20. A computer program product as in claim 19 wherein the plurality of classifiers
2 comprises classification trees, discriminant functions, regression trees, support vector
3 machines, neural nets and hidden Markov models.

1 21. A computer program product as in claim 20 further comprising computer code
2 means for extracting chart data from images, the extracted chart data being presented to
3 the user for selecting images for viewing.

1 22. A computer program product as in claim 21 further comprising:
2 computer code means for supplementing said extracted chart data with
3 supplemental data; and
4 computer code means for generating a chart representative of supplemented
5 chart data, said generated chart being provided responsive to said
6 query.

1 23. A computer program product as in claim 22 wherein the image types include
2 column charts, bar charts, line charts, pie charts, scatter charts, area charts, surface
3 charts, and three-dimensional charts.

1 24. A computer program product as in claim 23 wherein image features include
2 horizontal lines, vertical lines, percentage of white area, circular arcs and text.

1 25. A computer program product as in claim 24 further including an image
2 identification trainer comprising:
3 computer code means for providing a plurality of pre-classified training
4 images;
5 computer code means for classifying training images according to extracted
6 image features;
7 computer code means for selecting a particular one of said plurality of
8 classifiers as being an optimum classifier for a corresponding group
9 of training images;
10 computer code means for collecting a plurality of images available from
11 remotely connected computers;
12 computer code means for indexing said collected images; and
13 computer code means for providing an index of said collected images for
14 interrogation by users.